344–345 Nucleic Acids Research, 2001, Vol. 29, No. 1

ICB database: the *gyrB* database for identification and classification of bacteria

Kanako Watanabe, James S. Nelson, Shigeaki Harayama and Hiroaki Kasai*

Marine Biotechnology Institute, Kamaishi Laboratories 3-75-1 Heita, Kamaishi, Iwate 026-0001, Japan

Received September 5, 2000; Revised and Accepted November 1, 2000

ABSTRACT

The Identification and Classification of Bacteria (ICB) database (http:/www.mbio.co.jp/icb) contains currently available information about the DNA gyrase subunit B (*gyrB*) gene in bacteria. The database is designed to provide the scientific community with a reference point for using *gyrB* as an evolutionary and taxonomic marker. Nucleic and amino acid sequence data are currently available for over 850 strains, along with alignments at several different taxonomic levels and an exhaustive review of primer selection and background information.

INTRODUCTION

Small subunit ribosomal RNA (SSU rRNA) is currently the molecule of choice for bacterial molecular systematics, and this large body of data is supported by several dedicated databases (1,2). However, certain protein-coding genes may have select advantages over rRNA data. Higher levels of sequence variation (3) allow differentiation of closely related strains while the ability to translate DNA to protein sequences permits phylogenetic analysis of distantly related strains and more accurate sequence alignment (4). The Identification and Classification of Bacteria (ICB) database was conceived to provide a focal point dedicated to bacterial identification and classification using protein-coding genes. The database concentrates on providing extensive resources to support the sequence data, and we have initially specifically focused on DNA gyrase subunit B (*gyrB*), a type II topoisomerase found in bacteria that is capable of introducing negative supercoils into a relaxed closed circular DNA molecule (5). Since 1995, when universal primers for the gene became available (6), several publications have suggested that *gyrB* is a suitable gene for bacterial phylogeny, possessing essential attributes such as limited horizontal transmission and presence in all bacterial groups. Over the last 5 years, a considerable volume of *gyrB* sequence data has been generated and the ICB database aims to make these data accessible, in a practical format, to the scientific community.

DATA SOURCES

Much of the sequence data in the ICB database is drawn from the major sequence repositories [GenBank (7), EMBL Data Library (8) and DDBJ (9)]. Additional sequence data are deposited directly by us. The database contains more than 850 sequences of *gyrB* and its paralogue *parE*. PCR amplification and sequencing primer data is collected from references or provided by MBI.

DATABASE CONTENT

Initially, the ICB database was a searchable and BLASTable (10) collection of primarily *gyrB* sequences, with particular attention being paid to taxonomic status of the represented bacteria (11). The database has now been expanded to provide users with key resources for more practical utilisation of these data. We have concentrated on three main areas.

Alignments are now available for *gyrB* data at several taxonomic levels. Alignments of amino acid data are presented at the domain level and for each class of bacteria, and amino acid and nucleic acid alignments are available for each genus. All alignments are generated from amino acid data and, for nucleic acid alignments, reverse-translated. Alignments were generated by ClustalX 1.8 (12) and adjusted manually.

An extensive review of *gyrB* amplification and sequencing primers has been completed, and we have included a list of primers for the amplification of *gyrB*. It also includes several unpublished primers. User selection of suitable primers is now facilitated by (i) an overview map of primer positions on the *gyrB* gene of *Escherichia coli*; (ii) alignments of primers against specific bacterial groups; (iii) a review of groups for which primer combinations have been successfully used; and (iv) primer specificity for *gyrB* versus its related paralogue *parE*.

Protocol and reference background information has been produced to assist users in practical aspects of data generation and analysis.

DATABASE ACCESS AND USAGE

The ICB database is housed at the Marine Biotechnology Institute at Kamaishi, Japan and can be accessed at http://www.mbio.co.jp/icb. Data can be accessed in five different modes. Keyword searches return records based on taxon, gene and ICB reference number. The taxonomic browser permits location of data for specific bacterial classes, subclasses and genera. Alignments are also sorted by taxa and can be viewed in html or downloaded in FASTA format. BLAST analysis (10) queries the complete database. Protocols and primer resources are contained in static html files linked to the home

*To whom correspondence should be addressed. Tel: +81 193 26 6538; Fax: +81 193 26 6584; Email: hiroaki.kasai@kamaishi.mbio.co.jp

page. Users of the ICB database are asked to cite this article in their publications.

REFERENCES

1. Maidak,B.L., Cole,J.R., Lilburn,T.G., Parker,C.T., Saxman,P.R., Stredwick,J.M., Garrity,G.M., Li,B., Olsen,G.J., Pramanik,S., Schmidt,T.M. and Tiedje,J.M. (2000) The RDP (Ribosomal Database Project) continues. *Nucleic Acids Res.*, **28**, 173–174. Updated article in this issue: *Nucleic Acids Res.* (2001), **29**, 173–174.
2. Van de Peer,Y., De Rijk,P., Wuyts,J., Winkelmans,T. and De Wachter,R. (2000) The European Small Subunit Ribosomal RNA database. *Nucleic Acids Res.*, **28**, 175–176.
3. Ochman,H., and Wilson,A.C. (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.*, **26**, 74–86.
4. Gupta,R.S. (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.*, **62**, 1435–1491.
5. Khodursky,A.B., Peter,B.J., Schmid,M.B., DeRisi,J., Botstein,D., Brown,P.O. and Cozzarelli,N.R. (2000) Analysis of topoisomerase function in bacterial replication fork movement: Use of DNA microarrays. *Proc. Natl Acad. Sci. USA*, **97**, 9419–9424.
6. Yamamoto,S. and Harayama,S. (1995) PCR amplification and direct sequencing of *gyrB* with universal primers and its application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl. Env. Microbiol.*, **61**, 1104–1109.
7. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.
8. Stoesser,G., Baker,W., van den Broek,A.E., Camon,E., Hingamp,P., Sterk,P. and Tuli,M.A. (2000) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **28**, 19–23. Updated article in this issue: *Nucleic Acids Res.* (2001), **29**, 17–21.
9. Taneto,Y., Miyazaki,S., Ota,M., Sugawara,H. and Gojobori,T. (2000) DNA Data Bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucleic Acids Res.*, **28**, 24–26.
10. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
11. Kasai,H., Watanabe,K., Gasteiger,E., Bairoch,A., Isono,K., Yamamoto,S., and Harayama,S. (1998) Construction of the *gyrB* database for the identification and classification of bacteria. In Miyano,S. and Takagi,T. (eds), *Genome Informatics 1998*. Universal Academic Press, Tokyo, pp. 13–21.
12. Jeanmougin,F., Thompson,J.D., Gouy,M., Higgins,D.G. and Gibson,T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.*, **23**, 403–405.